采用异常检测的技术机会识别方法研究

翟东升 郭 程 张 杰 李登杰
(北京工业大学经济与管理学院 北京 100124)

摘要:【目的】为探索一种准确而及时地识别技术机会的方法，提出一种基于异常检测技术识别技术机会的框架。【方法】通过构建相似度矩阵进行多维尺度分析，基于多种异常点检测算法识别出潜在技术机会专利，结合 TRIZ 的技术系统进化法则从潜在技术机会专利中挖掘出技术机会。【结果】获取激光光刻技术 2000 年–2015 年的德温特专利数据，对该领域不同阶段的技术机会进行识别，分析结果显示，通过此框架识别出的前两个阶段的技术机会就是下一阶段的主流技术，同时，改进后的极紫外光刻技术可能成为下一代激光光刻领域的主流技术之一。【局限】利用 TRIZ 判定技术机会存在一定的主观性，识别准确度有待进一步提高。【结论】基于异常检测的技术机会识别方法可以有效地识别出技术机会，有助于提高识别技术机会的及时性。

关键词: 技术机会 异常检测 专利 TRIZ
分类号: G306.0 G353

1 引 言

在科学技术飞速发展的今天，创新能力已经成为衡量企业竞争力的一个重要因素。而作为技术创新的重要基础，技术机会越来越受到企业乃至整个行业的重视，如何尽可能早地发现技术机会，把握正确的技术方向，往往决定了企业未来发展的成败。因此，以技术机会理论为基础，从数据中挖掘出可能反映未来技术方向的技术机会，可以为企业的技术创新活动提供方向，具有重要的意义。

专利文献作为科技创新成果的重要载体和表现形式，其内容蕴涵了技术创新的前沿信息。据世界知识产权组织统计，世界上的新技术、新发明，有 90%至 95%记录在专利文献中，而专利公报中约有 80%的内容未曾刊载于其他专业期刊或学术论文等技术文献资料中[1]。由于专利数据具有易获取、数据相对完整等特点，故常用专利数据测度一定时期内的技术创新状况与技术发展水平。因此，本文选择专利作为研究的切入点，以专利作为研究对象探讨技术机会的问题。本研究运用多种异常检测算法，结合 TRIZ 理论中的技术系统进化法则识别潜在的技术机会，这对于技术机会识别是一种新的尝试。

2 国内外研究现状

国外学者对于技术机会的研究起步较早。1974 年，美国斯坦福大学的 Schwartz 教授提出技术创新机会(Technology Innovation Opportunity)的概念，并将其表述为能够带动某产业领域创新活动或使产业研发方向发生转折的关键技术[2]。20 世纪 90 年代，美国佐治亚理工学院教授 Porter 等提出“技术机会”一词，其领导的技术政策与评价中心(Technology Policy and Assessment Center)联合一些企业及组织开发技术机会分析(TOA)软件，基于数据并辅以计算机相关技术开展信息挖掘工作[3]。1996 年，汉城国立大学工业工程部的专家 Yoon 等进行了一项有关技术机会理论方法的研究，主要介绍了一种基于关键词的形态分析方法[4]。近几年，一些国外学者利用不同的方法开展了技术机会识别的研究。2009 年，Lee 等构造基于关键词的专利地图，将地图中的空白点视为可能的潜在技术机会，进行技术创新活动的识别[5]。Yoon 等提出一种基于 SAO 结构识别技术

通讯作者：郭程，ORCID：0000-0003-1205-0948，E-mail：guochengchn@163.com。

机会的方法，通过 SAO 结构量化专利的相似程度，使用异常检测算法识别可能代表技术机会的离群专利[6]。Lee 等提出一种两阶段专利分析方法，先基于专家构建技术属性–应用矩阵，识别基本的技术机会，然后抽取专利的 A-O 结构，挖掘出未被专家识别的属性或者应用，识别出新的技术机会[7]。Yoon 等结合形态分析和本文挖掘，基于现有技术进行技术机会识别[8]。Yoon 等则从已存在的技术或者产品的角度出发，从专利中获取产品、技术和功能的结构信息，利用技术或者产品之间的语义功能相似度识别潜在技术机会[9]。

国内学术界较早对技术机会开展研究的是中国管理科学院的李保明教授，他认为技术机会是指企业提供具有较大应用价值的新技术的可能性，并从技术、经济两个角度对其加以概括[10]。李辉等结合技术创新活动的发展方向，认为技术机会包括两个方面：以前未出现的新技术和对已有技术的改良、发展和创新[11]。黄鲁成等认为技术机会通常会在一个长期存在的重大技术障碍被克服之后出现，主要表现为新产品的涌现及创新集群的形成等[12]。此外，张妍则从哲学视角出发，对技术机会的来源、定义、发现等多个方面进行阐述[13]。近几年，国内学者尝试了多种方法对技术机会进行预测。吕一博等将 IPC 作为技术机会关键词，构建技术机会可视化识别图谱，基于图谱得出技术机会的所在领域[14]。潘东华等借鉴技术预测和知识发现的相关理论，建立关键技术、技术前沿和技术趋势三个视角的分析框架，并且通过共词分析、突现词分析和聚类时序图谱进行分析[15]。马婷婷等从系统的角度出发，基于技术传递系统，建立了一个从技术分析、竞争环境分析到潜在市场分析的三层分析架构[16]。汪雪锋等提出一种融合数据挖掘和形态分析的创新导图构建方法，通过“鱼骨图”和“云线图”对形态矩阵进行解读，绘制创新导图[17]。任智军等通过构建基于数据挖掘的技术机会发现模型，对新兴技术及技术成熟度进行预测[18]。郭俊芳等基于文本挖掘和 SAO 链语义分析，提出一种新型的技术形态识别方法，考虑了关键概念间的关系，更准确地构建了技术形态结构[19]。总体来说，现有关于技术机会的研究主要从已有一定数量的技术或者产品出发，对未来的技术机会做出预测。本文认为，对于技术机会的识别，不仅要关注技术的发展现状及发展轨迹，同时，应该重视技术机会识别的及时性，在技术机会形成的早期对其进行发掘。因此，本文利用异常检测和 TRIZ 的技术系统进化法则进行技术机会的识别，提高预测技术机会的及时性。

3 基于异常检测的技术机会识别

3.1 异常检测理论

异常检测的目标是发现与大部分其他对象不同的对象。通常，异常对象被称为离群点(Outlier)，因为在数据的散布图中，它们远离其他数据点。根据 Hawkins 的异常点定义，异常点是一个观测值，它与其他观测值的差别如此之大，以至于怀疑它是由不同的机制产生的[20]。而技术机会是在某技术领域中新发现的、以前未曾出现过的技术，或者是对现有技术的发展、改进和创新，它和现有的主流技术有很大的不同，因此，本文认为，在一定时期内某领域的技术机会存在于异常点中，可以通过异常检测的方法找出与主流技术有着明显不同的专利，在这些专利中，就有可能出现促进技术创新的专利，从而得出技术机会。

异常检测有三种基本方法：非监督的、监督的和半监督的。它们的主要区别是类标号(异常或正常)可以利用的程度[21]。由于技术机会事先是未知的，因此本文采用非监督的异常检测算法。

本文采用三种异常检测算法，即基于局部异常因子(LOF)的异常检测算法，基于 K 均值(K-means)的异常检测算法和基于聚类(DBSCAN)的异常检测算法。由于专利的分布特征是未知的，并且不同领域的专利的分布也是不一样的，因此，需要不同的异常检测算法识别具有不同特征的专利分布。K 均值算法简单并且可以用于各种数据类型，但需要事先确定类的个数，而在一般情况下，类的个数事先很难确定，而且它不能处理非球形簇、不同尺寸和不同密度的簇。LOF 和 DBSCAN 算法都是基于密度的聚类算法，它们能够处理任意形状和大小的簇，克服了 K 均值算法的缺点，然而，当簇的密度变化太大时，也会对聚类效果产生影响。

判断一个专利是否为异常点，需要有对此专利异常程度的评估，这种评估称为离群点得分。一个专利的离群点得分越高，说明此专利更有可能成为异常点。不同异常检测算法的离群点得分的计算方式是不同的。基于 K 均值的异常检测算法的离群点得分由到它的 K 个最近邻的平均距离决定[22]，其计算公式[23]为:

$$Outlier_Score_{k均值}(x)=\frac{\sum_{i=1}^{k}dist(x,y_i)}{k} \quad (1)$$

其中，$dist(x,y_i)$代表专利x与专利y_i之间的距离，专利y_i是距离专利x前K个最近邻专利中的一个。

基于LOF的异常检测算法的离群点得分由局部可达密度(Local Reachable Distance)决定，其计算公式[24]为：

$$Ird_{MinPts}(p)=\frac{1}{\frac{\sum_{o\in N_{MinPts(p)}}reach-dist_{MinPts}(p,o)}{|N_{MinPts}(p)|}} \quad (2)$$

专利p的局部可达密度是专利p与其MinPts-邻域的平均可达距离的倒数。其中，MinPts是用户指定的一个参数，$reach\text{-}dist_{MinPts}(p,o)$代表专利p与专利o的可达距离，而可达距离由专利o的k-距离与p到o的距离的最大值决定。$|N_{MinPts}(p)|$是p的邻域中邻居的数目。专利p的离群点得分定义[24]为：

$$Outlier_Score_{LOF}(p)=\frac{\sum_{o\in N_{MinPts(p)}}\frac{Ird_{MinPts}(o)}{Ird_{MinPts}(p)}}{|N_{MinPts}(p)|} \quad (3)$$

基于DBSCAN的异常检测算法的离群点得分是该专利的密度与它的最近邻专利的平均密度之比[25]。其计算公式[23]为：

$$density(x,k)=\left(\frac{\sum_{y\in N(x,k)}distance(x,y)}{|N(x,k)|}\right)^{-1} \quad (4)$$

$$Outlier_Score_{DB}(x,k)=\frac{density(x,k)}{\sum_{y\in N(x,k)}\frac{density(y,k)}{|N(x,k)|}} \quad (5)$$

其中，$density(x,k)$是专利x的密度，$|N(x,k)|$代表专利x的k-最近邻集合的大小；离群点得分越高，则说明专利x的异常程度越大。

3.2　技术机会识别模型的框架

本文基于专利文献数据，构建专利–关键词矩阵，计算专利之间的相似度；对专利相似度矩阵进行多维尺度分析并可视化，得到专利基于相似距离的二维坐标；使用多种异常检测算法组合的方法，识别出异常专利点，结合技术系统进化法则从中识别出潜在技术机会，构建技术机会识别模型，如图1所示：

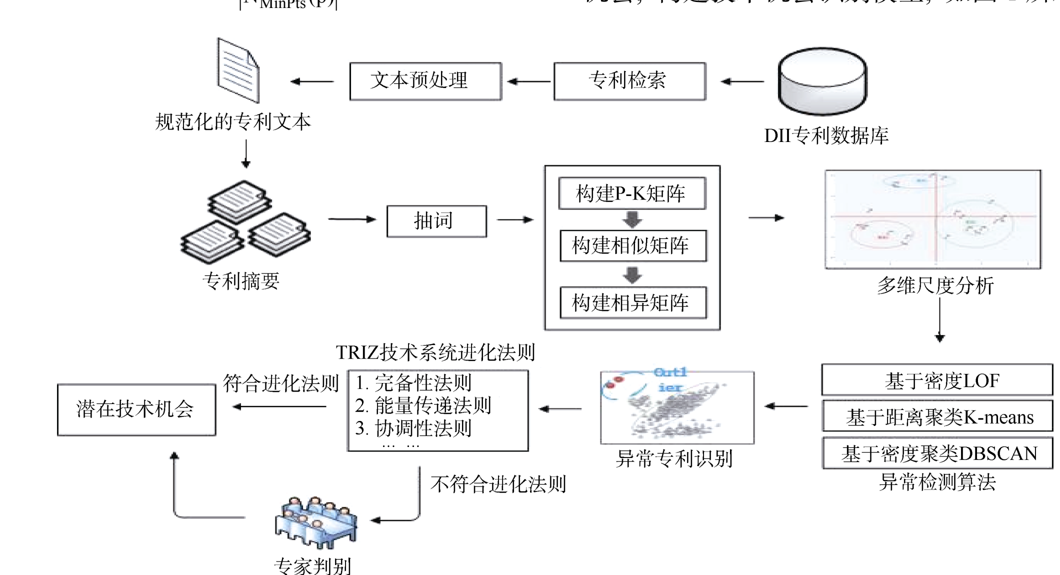


图 1　模型框架

(1) 专利相似矩阵的构建

本文从专利摘要部分抽取相应的关键词，构建存在映射关系的专利–关键词矩阵(P-K矩阵)，将专利文本表示为空间向量的形式，从而计算专利的相似度。

向量空间模型(Vector Space Modal)的基本思想是将文本表示成一个由特征项构成的向量，特性项通常是一个关键词，词与词之间没有相关性[26]。将专利文本通过向量空间模型的形式来表示，需要进行分词、停用词处理和关键词选取这三个主要步骤。由于本文所涉及的数据源都为英文专利文本，因此分词处理相对简单，根据空格、符号和段落对专利文本进行分词，得到词或词组；完成分词处理后，需要进行停用词处理，如“a”、“and”、“then”、“of”这类词出现的频率很高，会对文本的特征表示产生很大的干扰，因此需要对其进行过滤；

关键词的选择是构建向量过程中最重要的部分，然而，仅仅通过上述文本挖掘的方法抽取出来的关键词可能无法准确地描述技术特征，因此，在筛选的过程中，专家的判断也发挥着十分重要的作用。关键词被抽取之后，合适的关键词就会通过专家判断留下来，而无法反映专利特征的关键词会被丢弃。关键词向量如下所示：

$$v_i = (tf_{i1}, tf_{i2}, tf_{i3}, \cdots, tf_{in}) \quad (6)$$

其中，v_i 代表第 i 个专利，tf_{in} 代表第 i 个专利中第 n 个关键词出现的次数。对于构建好的 P-K 矩阵，通过余弦定理计算专利之间的相似度，公式如下：

$$\cos\theta = \frac{v_1 \times v_2}{\|v_1\|\|v_2\|} \quad (7)$$

最后，基于专利的相似度，就可以将专利–关键词矩阵转换为专利–专利的相似度矩阵。

(2) 多维尺度分析

多维尺度分析法(MultiDimensional Scaling, MDS)的主要思想是将含有多个变量的复杂数据压缩到一个低维空间，通过计算变量之间的距离研究各变量之间的结构，它将一组个体间的相异数据经过 MDS 转换成空间构图，且保留原始数据的相对关系[27]。

因此，使用 MDS 算法绘制专利相似地图。目前，有许多 MDS 算法，如 PREFSCAL，PROXCAL 和 ALSCAL。本文利用 ALSCAL 进行多维尺度分析，通常使用应力值 V(0≤V≤1)和拟合指数 RSQ 作为衡量 MDS 结果的指标。通常 V<0.2 是可接受的，RSQ 值在 0.6 以上是可接受的，具体判定方式如表 1 所示：

表 1　MDS 分析结果评价表

应力值(V)容差范围	MDS 分析效果
0.4≤V≤1	很差
0.2≤V<0.4	差
0.1≤V<0.2	一般
0.05≤V<0.1	好
0.025≤V<0.05	很好
0≤V<0.025	完美

因此，在使用 MDS 分析时，应当将应力值容差范围控制在 0.1 以下，使 MDS 分析的结果达到好或者很好。

本文基于专利相似度矩阵进行多维尺度分析，得到相似专利地图，以及专利在二维系中的位置坐标。通过多维尺度分析，可以将专利之间相似性作为判断依据，也可以将专利基于相似距离在二维坐标中进行定位，为下文中的各个异常检测算法奠定了基础。

(3) 技术机会识别

分别用三种异常检测算法对专利进行异常点的识别，得到异常专利点。对于这些专利，还不能认为其是技术机会，因为其中可能会有不符合技术发展规律的噪声专利。因此，本文基于 TRIZ 理论中的技术系统进化法则对其进一步识别。该进化法则包括完备性法则、能量传递法则、协调性法则、提高理想度法则、动态性进化法则、子系统不均衡进化法则、向微观级进化法则、向超系统进化法则，它们指出了技术系统进化发展的规律和宏观的模式与方向，反映了技术系统发展过程中会经历的具体阶段和进化顺序[28]。

因此，考虑对异常专利的具体技术内容进行分析，结合 TRIZ 的技术系统进化法则识别技术机会。首先，异常检测算法会将专利进行聚类，得到不同大小的专利簇，对非异常点的专利簇进行分析，得到现阶段的主流技术；然后，对异常专利点进行分析，若某一个异常专利的技术不完全等同于主流技术，并且从主流技术向异常专利的技术方向进行演变的路径符合 TRIZ 中的任意一条进化法则，则可以认为它是潜在的技术机会；若不符合，则将此专利交由专家进行判定。

4　实验及结果分析

4.1　数据来源

以激光光刻技术为检索对象，从技术领域、技术手段和技术效果等方面构建检索表达式，通过国际专利分类表(IPC)对检索领域进行限制，基于查全率和召回率对检索结果进行评估，当查全率和召回率均高于 90%则认为检索结果符合要求，得到该技术领域的专利文献，这些专利数据包括专利号、申请日期、专利标题、摘要，权利要求等内容。

本研究以德温特数据库为数据源，检索表达式为 TS=(("i-line" or "Argon fluoride" or AF or "Krypton fluoride" or KF or "extreme ultraviolet" or Photolithography) and semiconductor* and ("laser" or pulse*)) and IP=(G* or C* or H*)，时间范围限定为 2000 年 1 月 1 日到 2015 年 12 月 31 日，在德温特数据库检索并筛选相关专利 527 件，通过相关文献的查阅对该技术领域的技术发展情况进行分析，将其划分为三个时间段：

(1) 2000-2004 年专利共计 72 件；

(2) 2005-2009 年专利共计 191 件;

(3) 2010-2015 年专利共计 264 件。

各个时间段专利分布情况及技术发展趋势如表 2 所示:

表 2　激光光刻技术在半导体制造业中的发展趋势

阶段	专利数量	普遍使用技术	下一代技术
1	72	i-line&KrF	ArF
2	191	KrF&ArF	EUV
3	264	ArF&EUV	

4.2　专利相似度的计算及可视化

为了计算专利文本之间的相似度, 首先需要构建向量空间模型, 将专利文本表示为向量的形式, 即构建 P-K(专利–关键词)矩阵。通过对专利摘要文本进行分词、停用词处理, 得到相应的关键词; 然后基于专家判断, 筛选出能够代表该技术特征的关键词, 去除不相关的词汇, 得到最终的特征项, 构建专利文本的向量空间模型。表 3 为部分基于激光光刻技术领域专利文本构建的 P-K 矩阵。

表 3　2000 年–2003 年专利–关键词矩阵(部分)

	light source	irradiation	Photo resistfilm	resin	wave length
EP1113005	1	1			1
EP1126320			1		
EP1167349	2			2	1
EP1184723			1		
EP1343052	1		1	1	1

针对构建好的 P-K 矩阵, 基于余弦定理计算专利文本之间的相似度并得到专利相似度矩阵。表 4 为部分激光光刻技术专利之间的相似度对称矩阵。

表 4　2000 年–2003 年专利相似度对称矩阵(部分)

	EP1113005	EP1126320	EP1167349	EP1184723	EP1199603
EP1113005	1	0.211850857	0.270765	0	0.109985
EP1126320	0.211850857	1	0.222277	0	0.060193
EP1167349	0.270765181	0.222277112	1	0	0.246183
EP1184723	0	0	0	1	0
EP1199603	0.109985336	0.060192927	0.246183	0	1

采用 ALSCAL 算法进行多维尺度分析, 因此要将相似矩阵转化为相异矩阵(即用 1 减去相似矩阵)。将高维的专利信息映射到二维平面, 图 2 是三个阶段专利分布的散点图。

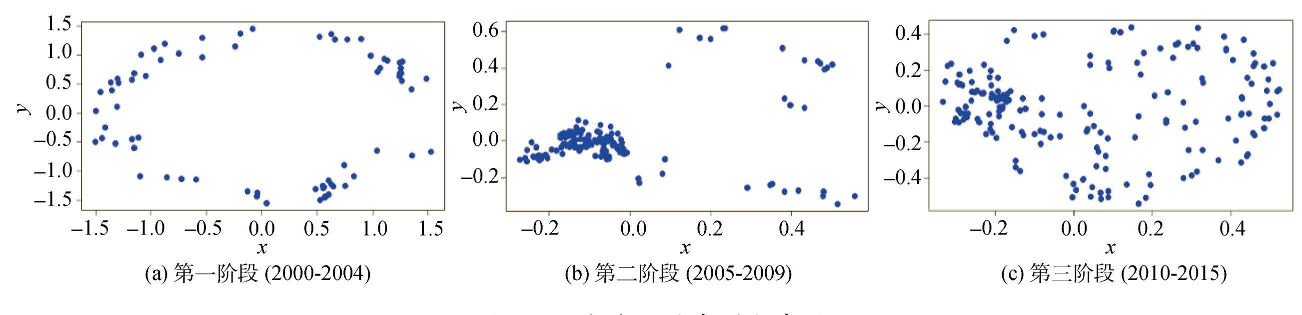


图 2　三个阶段的专利分布图

4.3　技术机会识别

对于三个阶段的专利, 分别使用基于 LOF 的异常检测、基于 K-means 的异常检测和基于 DBSCAN 的异常检测三种方法进行异常专利的识别, 结果如图 3 所示; 然后, 基于 TRIZ 的技术系统进化法则对异常专利的内容进行分析, 识别出技术机会。

对第一阶段的 72 条专利进行异常检测, 基于 LOF 的异常检测结果如图 3(a)所示, 红色圈内的专利为异常点, 即为异常专利, 检测出的异常专利为 JP2003147474, US2003186524, WO2003014830, US2003027349, US2002001957, US2003087188; 基于 K-means 的异常检测结果如图 3(b)所示, 不同颜色的点代表不同的类, 在此次聚类中, 设定 K=4, 即将激光光刻技术的相关专利聚为 4 类, 分别用黑色、绿色、蓝色和红色的专

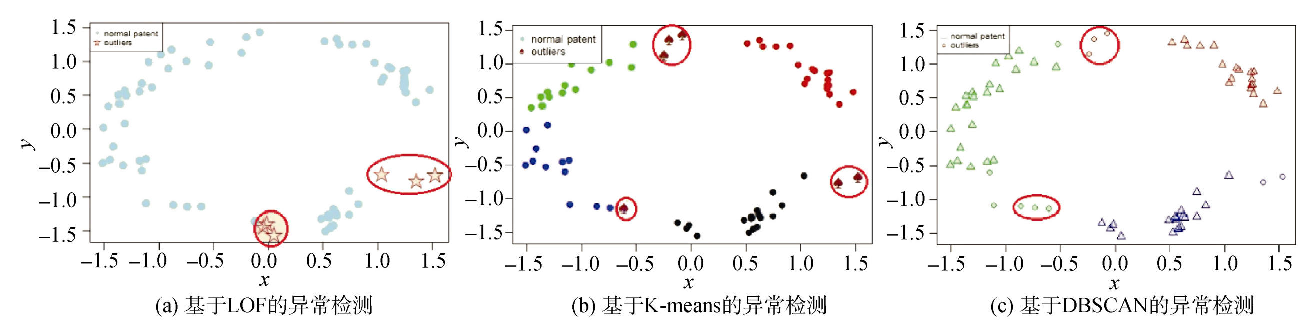


图 3　第一阶段的异常检测结果

利点代表，红色圈内的点代表偏离簇类中心的异常点，即为发现的异常专利点：JP2003147474，US2003186524，US2003107720，JP2000063441，US2002086222，US2002113035；通过 DBSCAN 算法对专利进行异常检测，结果如图 3(c)所示，在设定半径参数为 1，密度阈值 MinPts 为 15 的情况下，相关专利被聚成三类，分别以不同的颜色和形状进行标注。同时，三角形的点为聚类的核心点，而圆点(红色圈内)则为聚类的边缘点，在这里认为它是偏离聚类中心的异常点，这些检测出的异常专利点为：JP2003147474，US2003186524，US2003107720，JP2000063441，US2002086222，US2002113035。对以上三种异常检测算法的结果进行汇总，结果如表 5 所示：

表 5　第一阶段异常专利识别结果

LOF	K-means	DBSCAN
JP2003147474	JP2003147474	JP2003147474
US2003186524	US2003186524	US2003186524
WO2003014830	US2003107720	US2003107720
US2003027349	JP2000063441	JP2000063441
US2003000920	US2002086222	US2003000920
US2003087188	US2003000920	US2002113035

由表 5 可见，三种异常检测算法共识别出 10 个异常专利，通过分析这些专利的详细内容，发现专利 US2003000920 提出一种新的刻蚀方法，能够防止光致抗蚀剂图案的变形，改善光刻的可靠性，并且通过注入氩元素的电子束的方式硬化光致抗蚀剂，提高光刻胶图案的抗蚀性。而影响光刻质量的主要因素和难点之一就是刻蚀方法，该专利在刻蚀方法中有新的突破，总体上符合子系统的不均衡进化法则，即系统的进化速度受最慢的子系统制约；同时，该专利提出使用 ArF 准分子激光器，由于光刻系统的分辨率与激光的曝光波长成反比，氩的曝光波长相比于之前主流曝光技术(KrF)有更短的波长，有利于提高分辨率，符合提高理想度进化法则，即技术系统会朝着不断提高系统理想度的方向进化。另外，专利 US2003087188 也提出了使用能发出波长为 193 纳米的激光的 ArF 准分子激光器，提高分辨率，也符合提高理想度进化法则。而其他异常专利并没有按照 TRIZ 的技术进化法则的路径发展，并且经过专家判定，不认为它们是技术机会。

第一阶段的分析可以看出，未来光刻领域的技术会向使用波长更短的 ArF 准分子激光器的方向发展。同样，对第二阶段的专利进行异常检测，结果如图 4 所示：

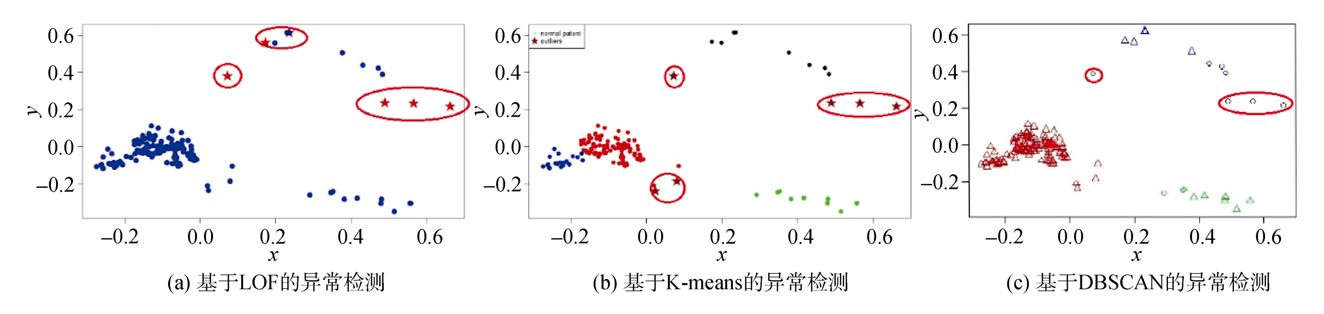


图 4　第二阶段的异常检测结果

图 4 中红色圈内的标出的点即为异常专利点。对三种异常检测算法的结果进行汇总，如表 6 所示：

表 6　第二阶段异常专利识别结果

LOF	K-means	DBSCAN
US2007123623	US2007123623	
JP2007220949	JP2007220949	US2007123623
EP1843201	EP1843201	JP2007220949
WO2007088862	WO2007088862	EP1843201
JP2004231858	EP1693709	WO2007088862
JP2005003863	JP.2005203649	

第二阶段共识别出了 8 个异常专利。通过分析这 8 条专利的详细内容，发现专利 JP2007220949 提出了一种用于光刻的极紫外光源装置，有一个激光束发生器，可以产生激光束来电离氢。极紫外线拥有更加短的波长，相对于之前的主流技术中使用的 ArF 准分子激光器，波长从 193 纳米下降到了 13.5 纳米，可以明显提高分辨率，符合提高理想度进化法则。同时，专利 EP1693709 提出一种抗蚀图案增厚材料，适用于形成微小的空间图案，突破曝光极限；并且抗蚀图案形成过程中包括用紫外线或者电离氢照射抗蚀图案的整个表面。由于曝光波长小于 100 纳米时，只通过缩短波长的方法已经很难更进一步提高分辨率，而此专利通过增厚材料突破曝光极限，通过其他方面的发展来提高分辨率，使整个系统更加协调，符合子系统协调性进化法则。因此，将这两条专利视为潜在的技术机会。而其他异常专利经由专家判断，认为其并不是技术机会。

从第二阶段的分析可以看出，未来光刻领域的技术会向波长明显短于之前技术的 EUV(极紫外光刻)技术的方向发展。

最后，对第三阶段的专利进行异常检测，如图 5 所示，红色圈内标出的点即为异常专利。对三种异常检测算法的结果进行汇总，如表 7 所示。

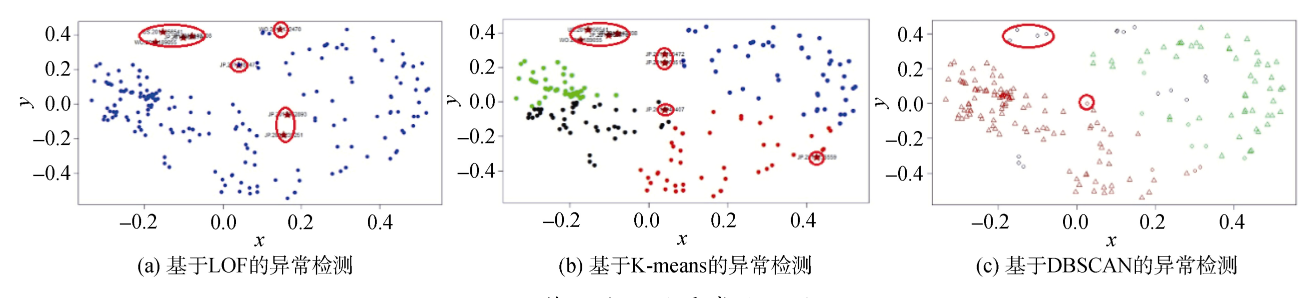


图 5　第三阶段的异常检测结果

表 7　第三阶段异常专利识别结果

LOF	K-means	DBSCAN
JP2015143208	JP2015143208	
JP2015063470	JP2015063470	
US2015056541	US2015056541	JP2015143208
WO2014189055	WO2014189055	JP2015063470
JP2012185472	JP2012185472	US2015056541
JP2013082893	JP2013003512	WO2014189055
JP2013235251	JP2014040407	
WO2015122470	JP2013136559	

第三阶段共识别出 11 条异常专利。通过分析这 11 条专利的详细内容，发现专利 US2015056541 提出使用极紫外光刻技术的空白掩膜用于制造半导体器件，它的优点是将对于掩膜的损害和污染降到最低，提高极紫外光刻掩膜的可靠性。虽然极紫外线光刻技术已经将波长降至 13.5 纳米，同时也伴随着一些其他的问题，制约着其分辨率的提高。而此专利提出掩膜的改进方案，从另一个方面提高光刻技术的精度，符合子系统协调性进化法则。因此，将它视为潜在技术机会。同时，将其他异常专利交由专家判定，专家认为，专利 WO2014189055 提出一种极紫外线激光发生装置的等离子(LPP)系统，具有目标生成控制单元，用于控制平均值压力控制器，有利于生成稳定的极紫外线激光。其也可能成为潜在的技术机会。

通过对第三阶段技术机会的识别，可以看出，未来光刻技术领域的技术主要还会是极紫外光刻技术，但是由于其波长明显缩短所带来的其他问题，通过提高其他方面性能的技术可能成为下一阶段的主流技术，如对掩膜技术的改造。

经过上述分析，通过异常检测方法和 TRIZ 的技术

系统进化法则，从第一阶段和第二阶段的专利中识别出可能成为技术机会的相关专利，并且与激光光刻技术的发展进程做对比，发现所识别出的技术确实是下一阶段的主流技术之一，因此，该方法在一定程度上可以用来识别潜在的技术机会，通过对当前专利中异常专利点的发现和判断，得出未来可能的发展方向。

同时，从上述结果中可以发现，三种异常检测算法对于技术机会的挖掘所产生的效果是不同的。在第一阶段的技术机会识别中，LOF 算法的表现要优于另外两种算法，发掘出了另外二者未识别出的异常专利，并且该异常专利确实是潜在的技术机会；在第二阶段的技术机会识别中，K-means 算法的表现更加优秀，多识别出一个潜在技术机会；而在第三阶段的识别中，三种算法都识别出两个潜在技术机会专利。由于三种算法在计算离群点得分时所用的方法不同，导致了识别效果的不同。K-means 算法很难处理非球形和不同大小的簇，并且在识别密度不相同的簇时效果很差，但是它可以发现不是明显分离的簇；而DBSCAN算法可以处理不同大小或者形状的簇，但对于密度不相同的簇的识别效果也不好；LOF 算法定义了相对密度的概念，因此有利于发现不同密度区域的异常点。第一阶段的专利分布并不均匀，并且所形成的簇都是不规则形状，因此 LOF 算法的效果相对较好；第二阶段的专利分布稀疏且不均匀，但是其中一个潜在技术机会专利并未与其他专利明显分离，因此 K-means 算法识别出了该异常点；而第三阶段的专利分布比较均匀，密度变化并不明显，因此三种异常点检测算法都识别出了潜在的技术机会。

由于不同领域的专利分布是不一样的，即使在同一领域，不同阶段的专利分布特点也会有一定的区别，因此，在进行异常专利的识别时，应该考虑多种异常检测算法，在不同特点的专利分布中更加全面地识别出异常专利。

5 结 语

本文基于专利文献，使用多种异常检测算法识别出异常专利，并且基于 TRIZ 的技术系统进化法则，对异常专利进一步分析，得到符合进化法则的异常专利，即潜在的技术机会。利用激光光刻技术领域的专利进行实证研究，对该领域发展的三个阶段的专利进行技术机会的识别，结果显示，从第一阶段和第二阶段中识别出的潜在技术机会确实是下一阶段的主流技术之一，说明该模型在一定程度上可以识别出技术机会；同时，对第三阶段的专利进行技术机会识别，发现激光光刻未来的技术可能会在极紫外光刻方向，并且对于掩膜技术的改进可能是一个技术机会。

同时，该模型存在一些可以改进的方面。在对 TRIZ 的技术系统进化法则的理解上，不同专业背景的人对同一个进化法则可能会有不同的理解，因此，对于模型中从异常专利到技术机会专利的识别过程中，会存在标准不一的问题。因此，应该建立一套规则，对 TRIZ 中的理论和专利中包含的技术进行更加深入的挖掘，实现 TRIZ 理论和专利之间的匹配。这也是未来需要研究和改进的地方。

参考文献：

[1] 包昌火，谢新洲. 竞争情报与企业竞争力[M]. 北京：华夏出版社，2001. (Bao Changhuo, Xie Xinzhou. Competitive Intelligence and Enterprise Competitive Power [M]. Beijing: Huaxia Publishing House, 2001.)

[2] Schwartz P. Technological Innovation Opportunities [J]. Computers and People, 1974, 23(5): 33.

[3] Porter A L, Detampel M J. Technology Opportunities Analysis [J]. Technological Forecasting & Social Change, 1995, 49(3): 237-255.

[4] Yoon B, Park Y. A Systematic Approach for Identifying Technology Opportunities: Keyword-based Morphology Analysis [J]. New England Journal of Medicine, 2005, 72(2): 145-160.

[5] Lee S, Yoon B, Park Y. An Approach to Discovering New Technology Opportunities: Keyword-based Patent Map Approach [J]. Technovation, 2009, 29(6): 481-497.

[6] Yoon J, Kim K. Detecting Signals of New Technological Opportunities Using Semantic Patent Analysis and Outlier Detection [J]. Scientometrics, 2012, 90(2): 445-461.

[7] Lee Y, Kim S Y, Song I, et al. Technology Opportunity Identification Customized to the Technological Capability of SMEs Through Two-stage Patent Analysis[J]. Scientometrics, 2014, 100(1): 227-244.

[8] Yoon B, Park I, Coh B Y. Exploring Technological Opportunities by Linking Technology and Products: Application of Morphology Analysis and Text Mining [J]. Technological Forecasting & Social Change, 2014, 86:

287-303.

[9] Yoon J, Park H, Seo W, et al. Technology Opportunity Discovery (TOD) from Existing Technologies and Products: A Function-based TOD Framework [J]. Technological Forecasting & Social Change, 2015, 100: 153-167.

[10] 李保明. 技术机会与技术创新的决策[J]. 科学管理研究, 1990, 8(5): 61-62. (Li Baoming. Decision of Technological Opportunities and Technological Innovation [J]. Researches in Science Management, 1990, 8(5): 61-62.)

[11] 李辉, 乔晓东. 基于科技文献的技术机会分析方法初探[J]. 情报杂志, 2007, 26(5): 74-76. (Li Hui, Qiao Xiaodong. The Technology Opportunity Analysis Based on Science & Technology Document [J]. Journal of Information, 2007, 26(5): 74-76.)

[12] 黄鲁成, 蔡爽. 基于专利的判断技术机会的方法及实证研究[J]. 科学学研究, 2010, 28(2): 215-220. (Huang Lucheng, Cai Shuang. An Empirical Study on the Evaluation of the Technical Opportunity Based on the Patents [J]. Studies in Science of Science, 2010, 28(2): 215-220.)

[13] 张妍. 市场创新活动中的技术机会[J]. 科技管理研究, 2011(3): 31-34. (Zhang Yan. Technological Opportunities in the Market Innovation Activities [J]. Science and Technology Management Research, 2011(3): 31-34.)

[14] 吕一博, 康宇航, 王淑娟. 基于共现分析的技术机会发现与可视化识别[J]. 科研管理, 2012, 33(4): 80-85. (Lv Yibo, Kang Yuhang, Wang Shujuan. Visualized Identification and Discovery of Technology Opportunities Based on Co-occurrence Analysis [J]. Science Research Management, 2012, 33(4): 80-85.)

[15] 潘东华, 徐珂珂. 基于共词分析的技术机会分析[J]. 科研管理, 2014, 35(4): 10-17. (Pan Donghua, Xu Keke. Technology Opportunity Analysis Based on Co-word Analysis [J]. Science Research Management, 2014, 35(4): 10-17.)

[16] 马婷婷, 汪雪锋, 朱东华, 等. 基于专利的技术机会分析方法研究[J]. 科学学研究, 2014, 32(3): 334-342. (Ma Tingting, Wang Xuefeng, Zhu Donghua, et al. Research on Technology Opportunity Analysis Based on Patents [J]. Studies in Science of Science, 2014, 32(3): 334-342.)

[17] 汪雪锋, 李兵, 许幸荣, 等. 基于形态分析法的创新导图构建及应用研究[J]. 科学学研究, 2014, 32(2): 178-183. (Wang Xuefeng, Li Bing, Xu Xingrong, et al. Research on Construction and Application of Innovation Mapping Based on Morphology Analysis [J]. Studies in Science of Science, 2014, 32(2): 178-183.)

[18] 任智军, 乔晓东, 徐硕, 等. 基于数据挖掘的技术机会发现模型研究[J]. 情报杂志, 2015, 34(6): 174-177. (Ren Zhijun, Qiao Xiaodong, Xu Shuo, et al. An Approach for Technology Opportunities Discovery Model Based on Data Mining [J]. Journal of Information, 2015, 34(6): 174-177.)

[19] 郭俊芳, 汪雪锋, 李乾瑞, 等. 一种新型的技术形态识别方法——基于 SAO 语义挖掘方法[J]. 科学学研究, 2016, 34(1): 13-21. (Guo Junfang, Wang Xuefeng, Li Qianrui, et al. A New Method to Identify Technology Morphology: SAO-based Semantic Analysis Approach [J]. Studies in Science of Science, 2016, 34(1): 13-21.)

[20] Duan L, Xu L, Liu Y, et al. Cluster-based Outlier Detection [J]. Annals of Operations Research, 2009, 168(1): 151-168.

[21] 张晓惠, 林柏钢. 基于特征选择和多分类支持向量机的异常检测[J]. 通信学报, 2009, 30(10): 68-73. (Zhang Xiaohui, Lin Bogang. Anomaly Detection Based on Feature Selection and Multi-class Support Vector Machines [J]. Journal on Communications, 2009, 30(10): 68-73.)

[22] Ramaswamy S, Rastogi R, Shim K. Efficient Algorithms for Mining Outliers from Large Data Sets [J]. ACM SIGMOD Record, 2000, 29(2): 427-438.

[23] 谭庞宁. 数据挖掘导论: 完整版[M]. 范明, 范宏建译. 第 2 版. 北京: 人民邮电出版社, 2011: 411-413. (Tan Pangning. Introduction to Data Mining [M]. Translated by Fan Ming, Fan Hongjian. The 2nd Edition. Beijing: The People's Posts and Telecommunications Press, 2011: 411-413.)

[24] Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying Density-based Local Outliers [J]. ACM SIGMOD Record, 2000, 29(2): 93-104.

[25] Ester M, Kriegel H P, Jiirg S, et al. A Density Based Algorithm for Discovering Clusters in Large Spatial Databases [C]. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. AAAI Press, 1996.

[26] 郭庆琳, 李艳梅, 唐琦. 基于 VSM 的文本相似度计算的研究[J]. 计算机应用研究, 2008, 25(11): 3256-3258. (Guo Qinglin, Li Yanmei, Tang Qi. Similarity Computing of Documents Based on VSM [J]. Application Research of Computers, 2008, 25(11): 3256-3258.)

[27] 储节旺, 闫士涛. 知识管理学科体系研究(下)——聚类分析和多维尺度分析[J]. 情报理论与实践, 2012, 35(3): 5-9. (Chu Jiewang, Yan Shitao. Research on the System of Knowledge Management — Cluster Analysis and Multi-dimensional Scaling Analysis [J]. Information Studies:

Theory & Application, 2012, 35(3): 5-9.)
[28] 程序，于海燕. TRIZ 理论在专利技术路线图制定中的应用研究[J]. 情报学报, 2012, 31(10): 1045-1051. (Cheng Xu, Yu Haiyan. Study on Application of TRIZ in Patent Technology Roadmapping [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(10): 1045-1051.)

作者贡献声明:

翟东升：提出研究选题及思路;
翟东升，郭程，张杰：设计并完善研究方案;
郭程，李登杰：采集、清洗数据，进行实验;
翟东升，郭程：论文起草;
翟东升：论文最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。
[1] 翟东升，郭程. phase1.txt. 第一阶段德温特专利数据.
[2] 翟东升，郭程. phase2.txt. 第二阶段德温特专利数据.
[3] 翟东升，郭程. phase3.txt. 第三阶段德温特专利数据.
[4] 郭程，李登杰. Similarity1.xlsx. 第一阶段专利相似矩阵.
[5] 郭程，李登杰. Similarity2.xlsx. 第二阶段专利相似矩阵.
[6] 郭程，李登杰. Similarity3.xlsx. 第三阶段专利相似矩阵.

收稿日期: 2016-05-29
收修改稿日期: 2016-08-07

Identifying Technology Opportunities with Anomaly Detection Technique

Zhai Dongsheng　Guo Cheng　Zhang Jie　Li Dengjie
(School of Economics and Management, Beijing University of Technology, Beijing 100024, China)

Abstract: [**Objective**] This paper proposes a framework to effectively identify technology opportunities with anomaly detection technique. [**Methods**] First, we constructed a similarity matrix and conducted multidimensional scaling analysis. Second, we identified potential technology opportunity from patents based on a variety of anomaly detection algorithm. Finally, we extracted the possible breakthroughs with the help of TRIZ's laws of technology system evolution. [**Results**] We analyzed patent data from the DII database and then identified technology opportunities in different phases of the laser lithography field. We found that technology opportunities identified by the proposed framework became mainstream technologies later. [**Limitations**] The objectiveness and accuracy of the new method needs to be improved. [**Conclusions**] The proposed framework based on anomaly detection could effectively identify technology opportunities.

Keywords: Technology opportunity　Anomaly detection　Patent　TRIZ